UNITED STATES OF AMERICA

APPLICATION FOR PATENT

FOR INVENTION OF

**DEVICES, SOFTWARES AND METHODS FOR
PRIORITIZING BETWEEN VOICE DATA PACKETS
FOR DISCARD DECISION PURPOSES**

Inventor(s):   Ning Mo
Carlos Laux
Geethgayathri Ramachandran
Chunyan Li

Application prepared by: Marger Johnson & McCollom
1030 S.W. Morrison Street
Portland, Oregon 97205 USA
tel: (503) 222-3613
fax: (503) 274-4622
www.techlaw.com

DEVICES, SOFTWARES AND METHODS FOR PRIORITIZING BETWEEN VOICE DATA PACKETS FOR DISCARD DECISION PURPOSES

CROSS REFERENCE TO RELATED APPLICATIONS

This document may be found to be related to U.S.A. Patent Application No. [SER.NO.OF.201], filed on January 3, 2002.

BACKGROUND OF THE INVENTION

1.   Field of the invention.

The present invention is related to the field of communications through networks, and more specifically to devices, softwares and methods for prioritizing between voice data packets for discard decision purposes.

2.   Description of the related art.

Networks, such as the internet, are increasingly used for communications. The Internet Protocol (IP) has been developed for communications through the internet.

Like other packet network technologies, the Internet Protocol (IP) was initially designed to transport data in a reliable and robust way. IP ensures that all data is delivered, even though in an asynchronous mode.

It was only later that networks started being used for transporting video data and voice data. The latter takes place using a Voice over Internet Protocol (VoIP). VoIP ensures that voice data are retransmitted in real time. If any are lost, they are not retransmitted. If any arrive too late, they are not incorporated in the playout.

Since the initial design of IP focused on transferring data asynchronously, it largely ignored considerations such as Quality of Service (QoS) for VoIP. Accordingly, as VoIP evolves, more and more efforts are being made to ensure an acceptable QoS over networks, such as IP networks. Schemes are developed, such as Resource Reservation Protocol (RSVP) and Class Based Queuing (CBQ). These schemes also address how to handle the difficult situations, such as congestion in a network, etc.

CBQ techniques define several logical queues in a memory buffer. Each queue is assigned a priority. As each incoming packet is received, it is sorted into one of these queues, where it awaits retransmission. Sorting is based on a number of criteria, such as the Type of Service (ToS) field of the packet, or a combination of source address,

5   destination address and port. Then a scheduler selects one of the queues based on its relative priority, and forwards packets from it to the interface physical queue, from where they are routed.

Since voice and video data packets have to be retransmitted at real time, they tend to be given a higher priority than data. And that is in networks that were initially

10   designed for data, instead of voice or video.

When there is a congestion in a network due to voice traffic, some of the packets carrying voice data are dropped. This means they are discarded, without being retransmitted, which guarantees they will not arrive for playout.

Discarding voice packets takes place randomly, depending only on the condition

15   of network congestion. The CBQ techniques do not differentiate between packets within a queue.

The speech is reconstructed for playout. Packets that got dropped along the way are typically reconstructed by interpolation from their neighboring packets that did arrive.

The problem is that, for reconstructing speech, some packets are more

20   perceptually important than others. But their relative importance is not accounted for in the discard decisions of a congested router. Accordingly, an important packet has an equal chance of being discarded as a less important packet. When important packets are discarded, interpolation aggravates the fact that they are missing. Thus the reconstructed voice can have a poor quality, even if the network is only mildly congested.

25

BRIEF SUMMARY OF THE INVENTION

The present invention overcomes these problems and limitations of the prior art.

Generally, the present invention provides devices, softwares and methods for prioritizing between voice data packets for discard decision purposes. A perceptual

30   importance of a voice data packet relative to the others is determined at encoding. The relative importance is represented in the packet as a comparative discardability code. If a

discard decision is made, it takes into account the comparative discardability code of the packet.

The invention offers the advantage that the determination of which packets of a stream should be discarded is made at the location of where encoding takes place. The determination of which packets are more important than the others is advantageously made according to the content of the encoded sound. Accordingly, when discarding takes place, the less important packets are discarded more often. Since the important packets survive dropping more frequently, the resulting reconstructed stream is affected minimally, even in the face of heavy network congestion.

The invention will become more readily apparent from the following Detailed Description, which proceeds with reference to the drawings, in which:

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a network diagram showing a voice gateway made according to an embodiment of the present invention establishing a telephone call connection via a router made according to an embodiment of the present invention through a network.

Fig. 2 is another network diagram showing an IP telephone made according to an embodiment of the present invention establishing a telephone call connection through the network of Fig. 1.

Fig. 3 is a block diagram of a network device made according to the invention.

Fig. 4 is a diagram of a data packet made, transmitted and received according to an embodiment of the present invention.

Fig. 5 is a flowchart illustrating a method according to an embodiment of the present invention.

Fig. 6 is a flowchart illustrating a method for performing a box of the flowchart of Fig. 5 according to an embodiment of the present invention.

Fig. 7 is a block diagram of another network device made according to the invention.

Fig. 8 is a flowchart illustrating a method according to yet another embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT(S)

As has been mentioned, the present invention provides devices, softwares and methods for prioritizing between voice data packets for discard decision purposes. The invention is now described in more detail.

5   Referring now to Fig. 1, a diagram is shown for network 100. Network 100 may be any packet switched communications network, such as the internet, a local area network (LAN), a metropolitan area network (MAN), an intranetwork of an organization, etc.

A telephone 110 is a common, circuit switched telephone. Its user makes a call to
10   a telephone 120, which is a packet switched telephone, also known as IP telephone. Since telephone 120 is accessible through network 100, a connection is established through network 100.

More particularly, telephone 110 first establishes a connection 114 with a voice gateway 130 in network 100. Voice gateway 130 establishes a packet switched
15   connection 134 with a router 140, and router 140 establishes a packet switched connection 144 with telephone 120, to complete the connection. In fact, there may also be other routers in the path, in addition to router 140.

Voice gateway 130 is made according to an embodiment of the present invention, as will be understood from the below. Voice gateway 130 transmits voice data packets to
20   router 140 along connection 214. At least one such packet 410 is shown.

Router 140 is also made according to an embodiment of the invention. It receives packets along connection 134, such as packet 410. Router 140 may experience congestion, in which case it may have to discard some of the voice data packets, without retransmitting them to IP telephone 120. Router 140 looks to the content of packets, such
25   as packet 410, to make a better educated decision as to which ones of the packets to discard.

The above is simply one of many possible configurations for the invention to work. If either voice gateway 130 or router 140 is not made according to the invention, retransmission will indeed happen, but the benefit of the invention will not be realized.
30   Particularly, if a router is not made according to the invention, it will not look to the packets, to determine which packets are more desirable to discard, as in the prior art.

Alternately, if a voice gateway is not made according to the invention, router 140 may look but not understand which packets are more desirable to discard.

Referring to Fig. 2, another configuration is shown, which has similarities with that of Fig. 1, so as to facilitate the present description. An IP telephone 210 made

5 according to the invention establishes a call with the same IP telephone 120 as in Fig. 1, through the same network 100. IP telephone 210 establishes a packet switched connection 214 with router 140. The remainder of the connection is the same.

IP telephone 210 is made according to an embodiment of the present invention, as will be understood from the below. IP telephone 210 transmits voice data packets to

10 router 140 along connection 214. At least one such packet 410 is shown.

Referring now to Fig. 3, a network device 300 made according to an embodiment of the invention is described in more detail. Device 300 may be a voice gateway such as voice gateway 130, an IP telephone such as IP telephone 210, etc.

Device 300 may be implemented by combining separate components.

15 Alternately, one or more of the components of device 300 may be implemented as an Application Specific Integrated Circuit (ASIC), etc.

Device 300 has a network interface 320 for interfacing with a network, such as network 100.

Device 300 also has a processor 340 coupled with network interface 320.

20 Processor 340 may include a codec 350 which is made from a voice encoder 360 and a voice decoder 370.

Processor 340 may be implemented as a Central Processing Unit (CPU), or any other equivalent way known in the art. In one embodiment, device 300 additionally includes a memory 380, on which a program 390 may reside. Functions of processor 340

25 may be controlled by program 390, as will become apparent from the below. Alternately, processor 340 may be implemented as a Digital Signal Processor (DSP), etc.

Referring to Fig. 4, a diagram is shown of a data packet 410. Packet 410 is made according to an embodiment of the present invention by device 300. Packet 410 is transmitted through network 100 along a packet switched connection, such as connection

30 134 of Fig. 1 or connection 214 of Fig. 2.

Packet 410 includes a payload 420 and a header 430. Payload 420 includes at least one encoded voice frame EVF of the telephone conversation. Frame EVF is made from data bits. Header 430 is interpreted by a retransmitting network device, to direct where packet 410 will be sent to.

5   Packet 410 includes a comparative discardability code CDC according to the invention. Code CDC indicates a discardability of frame EVF relative to frames in other packets (not shown).

Code CDC may be located anywhere in packet 410. It is highly preferred that code CDC be part of header 430. For example, header 430 may be a Real-Time

10   Transport Protocol (RTP) header, and code CDC may be part of an extension of RTP header 430.

Code CDC may be just one bit. By convention, the bit may be "1" to signify a higher discardability of a packet whose bit is "0".

As will be understood also from the below, code CDC does not determine for

15   certain whether packet 410 will be discarded or not. If there is no congestion in network 100, then packet 410 will probably not be discarded at all. If network 100 is congested, then a code of "1" will make it a more likely discard candidate, than a packet a code of "0".

This selection process at router 140 will result in salient portions of the speech

20   having a better chance of being transmitted, as opposed to non salient such portions. Reconstruction, therefore, will produce better sounding voice for the user of telephone 120.

The invention may be practiced if all the voice data packets are configured as packet 410, but that is not necessary. Only some voice data packets out of the entire

25   stream need have a CDC code. The remaining voice data packets may, by convention, be deemed more desirable or less than those with the CDC codes.

The present invention may be implemented by one or more devices that include logic circuitry. The device performs functions and/or methods as are described in this document. The logic circuitry may include a processor that may be programmable for a

30   general purpose, or dedicated, such as microcontroller, a microprocessor, a Digital Signal Processor (DSP), etc. For example, the device may be a digital computer like device,

such as a general-purpose computer selectively activated or reconfigured by a computer program stored in the computer. Alternately, the device may be implemented an Application Specific Integrated Circuit (ASIC), etc.

Moreover, the invention additionally provides methods, which are described below. The methods and algorithms presented herein are not necessarily inherently associated with any particular computer or other apparatus. Rather, various general-purpose machines may be used with programs in accordance with the teachings herein, or it may prove more convenient to construct more specialized apparatus to perform the required method steps. The required structure for a variety of these machines will become apparent from this description.

In all cases there should be borne in mind the distinction between the method of the invention itself and the method of operating a computing machine. The present invention relates both to methods in general, and also to steps for operating a computer and for processing electrical or other physical signals to generate other desired physical signals.

The invention additionally provides programs, and methods of operation of the programs. A program is generally defined as a group of steps leading to a desired result, due to their nature and their sequence. A program made according to an embodiment of the invention is most advantageously implemented as a program for a computing machine, such as a general-purpose computer, a special purpose computer, a microprocessor, etc.

The invention also provides storage media that, individually or in combination with others, have stored thereon instructions of a program made according to the invention. A storage medium according to the invention is a computer-readable medium, such as a memory, and is read by the computing machine mentioned above.

The steps or instructions of a program made according to an embodiment of the invention requires physical manipulations of physical quantities. Usually, though not necessarily, these quantities may be transferred, combined, compared, and otherwise manipulated or processed according to the instructions, and they may also be stored in a computer-readable medium. These quantities include, for example electrical, magnetic, and electromagnetic signals, and also states of matter that can be queried by such signals.

It is convenient at times, principally for reasons of common usage, to refer to these quantities as bits, data bits, samples, values, symbols, characters, images, terms, numbers, or the like. It should be borne in mind, however, that all of these and similar terms are associated with the appropriate physical quantities, and that these terms are merely

5   convenient labels applied to these physical quantities, individually or in groups.

This detailed description is presented largely in terms of flowcharts, display images, algorithms, and symbolic representations of operations of data bits within at least one computer readable medium, such as a memory. An economy is achieved in the present document in that a single set of flowcharts is used to describe both methods of the

10   invention, and programs according to the invention. Indeed, such descriptions and representations are the type of convenient labels used by those skilled in programming and/or the data processing arts to effectively convey the substance of their work to others skilled in the art. A person skilled in the art of programming may use these descriptions to readily generate specific instructions for implementing a program according to the

15   present invention.

Often, for the sake of convenience only, it is preferred to implement and describe a program as various interconnected distinct software modules or features, individually and collectively also known as software and softwares. This is not necessary, however, and there may be cases where modules are equivalently aggregated into a single program

20   with unclear boundaries. In any event, the software modules or features of the present invention may be implemented by themselves, or in combination with others. Even though it is said that the program may be stored in a computer-readable medium, it should be clear to a person skilled in the art that it need not be a single memory, or even a single machine. Various portions, modules or features of it may reside in separate

25   memories, or even separate machines. The separate machines may be connected directly, or through a network, such as a local access network (LAN), or a global network, such as the Internet.

It will be appreciated that some of these methods may include software steps which may be performed by different modules of an overall parts of a software

30   architecture. For example, data forwarding in a router may be performed in a data plane, which consults a local routing table. Collection of performance data may also be

performed in a data plane. The performance data may be processed in a control plane, which accordingly may update the local routing table, in addition to neighboring ones. A person skilled in the art will discern which step is best performed in which plane.

In the present case, methods of the invention are implemented by machine

5   operations. In other words, embodiments of programs of the invention are made such that they perform methods of the invention that are described in this document. These may be optionally performed in conjunction with one or more human operators performing some, but not all of them. As per the above, the users need not be collocated with each other, but each only with a machine that houses a portion of the program.

10   Alternately, some of these machines may operate automatically, without users and/or independently from each other.

Methods of the invention are now described. It will be appreciated that some of these methods may include software steps which may be performed by different modules of an overall parts of a software architecture. For example, data forwarding in a network

15   switch (such as a router or a bridge) may be performed in a data plane, which consults a local routing table. Collection of performance data may also be performed in a data plane. The performance data may be processed, and accordingly used in a control plane to update the local routing table, in addition to neighboring ones. A person skilled in the art will discern which step is best performed in which plane.

20   Referring now to Fig. 5, a flowchart 500 is used to illustrate a method according to an embodiment of the invention. The method of flowchart 500 may also be practiced by voice gateway 130, IP telephone 210, device 300, etc.

According to a box 510, voice signals are received.

According to a next box 520, the voice signals are grouped into data speech

25   frames.

According to a next box 530, the voice signals of at least some of the frames are analyzed. Each frame is thus classified according to a type or class of speech.

From studies of speech production process, human speech sounds can be classified into three distinct classes according to their production process.

Voiced sounds are produced by forcing air through the glottis with the tension of the vocal cords adjusted, so that they vibrate in a relaxation oscillation, thereby producing quasi-periodic pulses of air, which in turn excite the vocal trace.

Unvoiced sounds are generated by forming a constriction at some point in the vocal tract (usually toward the mouth end), and forcing air through the constriction at a high enough velocity to produce turbulence. This creates a broad-spectrum noise source to excite the vocal tract.

Polsive sounds result from making a complete closure (again, usually toward the front of the vocal tract), building up pressure behind the closure, and abruptly releasing it.

Studies also show that 60% of the time the caller doesn't talk (silence period) for a normal telephone conversation. Discontinuous transmission schemes (silence compression using Voice Activated Detection (VAD) and Comfort Noise Generation (CNG)] are commonly used to reduce bandwidth required for voice traffic. Even with advanced VAD algorithms, however, there has to be some hangover time of encoding silence to avoid backend clipping. So with VAD on, many of the transmitted voice packets are still silence frames.

According to a next box 540, a comparative discardability is determined for some of the data speech frames relative to others. The determination is made according to the perceptual importance of the speech frames relative to each other. The relative importance is determined from the type or class of speech, and from some empirical data. Namely, from the auditory perception point of view, the human brain is sensitive to transitions from one kind of sound to another, for example, from Voiced sound to Unvoiced sound, or Unvoiced to Voiced. In contrast, the human brain is not so sensitive to the missing gap in between one kind of continuous sound, because it is smart enough to interpolate the missing part if the gap is short.

According to a next box 550, the data speech frames are encapsulated into data packets, such as packet 410. These data packets include comparative discardability codes indicating the determined comparative discardability of the encapsulated data speech frames

According to a next box 560, the data packets are transmitted through a packet switched network, such as network 100.

Referring now to Fig. 6, a flowchart 600 is used to illustrate a preferred method for performing box 540 of Fig. 5. It will be appreciated that flowchart 600 exploits the known empirical data mentioned above.

According to a box 610, a present speech type variable PRES_TYPE is set to any one of possible types of speech. The preferred one is S, which is defined for silence.

According to a next box 613, a counter K of frames is set to 1. If this is not the first time box 620 is executed, then counter K is reset to 1.

According to a next box 616, a previous speech type variable PREV_TYPE is set as equal to the present speech type variable PRES_TYPE.

According to a next box 620, a next voice frame is input for consideration.

According to a next box 630, a speech type of the frame is determined, and stored as the present speech type variable PRES_TYPE. The speech type may be S (for silence), U (for unvoiced), V (for voiced) and P (for Polsive). Other groupings of the types are also possible.

According to a next box 640, it is inquired whether the present speech type variable PRES_TYPE encodes S for silence. If yes, then according to a next box 645, a code CDC for the frame is set to 1 (for discardable), and execution returns to box 620. In other words, a data packet encapsulating a frame of silence is assigned a high comparative discardability.

If not, then according to a next box 650, it is inquired whether the present speech type variable PRES_TYPE encodes the same as the previous type speech variable PREV_TYPE. If not, it means there is a transition in speech type. Then according to a next box 655, a code CDC for the frame is set to 0 (for non-discardable), and execution returns to box 620. In other words, a data packet encapsulating a frame that transitions from one type of speech to another is assigned a low comparative discardability.

If not, then according to a next box 660, it is optionally inquired whether the counter K has reached up to one less from a first preset number NSET. The number NSET may be the same for all the types of speech, or different between types. If yes, it means that there are a number of frames with the same type of speech in a row. Then according to an optional next box 665, a code CDC for the frame is set to 1 (for discardable), and execution returns to box 620. A good value for NSET is 5.

In other words, when a segment has the same type of speech for many frames, every 5th frame is assigned a higher comparative discardability. It may be dropped, and the deterioration in the perception of sound will be the least. Plus, the reconstructive interpolation at playout will rectify this type of missing packet with the best fidelity.

5   If at box 660 the counter K has not reached up to one less from number NSET, then according to a next box 670, a code CDC for the frame is set to 0 (for non-discardable).

According to a next box 680, the counter K is incremented by 1. Then execution returns to box 620.

10   Referring now to Fig. 7, a network device 700 made according to an embodiment of the invention is described in more detail. Device 700 may be a router such as router 140, a bridge, etc.

Device 700 may be implemented by combining separate components. Alternately, one or more of the components of device 700 may be implemented as an

15   Application Specific Integrated Circuit (ASIC), etc.

Device 700 has a network interface 720 for interfacing with a network, such as network 100.

Device 700 also has a processor 740 coupled with network interface 720. Processor 740 may be implemented as a Central Processing Unit (CPU), or any other

20   equivalent way known in the art. In one embodiment, device 700 additionally includes a memory 780, on which a program 790 may reside. Functions of processor 740 may be controlled by program 790, as will become apparent from the below. Alternately, processor740 may be implemented as a Digital Signal Processor (DSP), etc.

Memory 780 has a portion allocated as a waiting queue 796. Queues are also

25   called buffers. Received packets, such as packet 410 are stored in buffer 796 until retransmission.

Referring now to Fig. 8, a flowchart 800 is used to illustrate a method according to an embodiment of the invention. The method of flowchart 800 may also be practiced by router 140, network device 700, etc.

30   According to a box 810, voice data packets are received through a packet switched network such as network 100.

According to a next box 820, the received packets are stored in a queue, such as buffer 796.

According to a next box 830, some of the stored packets are retransmitted through the network from the queue.

5   According to a next box 840, it is inquired whether congestion is sensed in the network. If not, then execution returns to box 810.

If yes, then according to a next box 850, a comparative discardability code CDC of a specific stored packet relative to the others is extracted. The CDC is known to be related to preset types of speech.

10   According to an optional next box 860, a discarding probability is set in accordance with the comparative discardability code CDC, and the preset types of speech. If CDC is 1, the discarding probability is set higher than would be otherwise. If CDC is 0, the discarding probability is set lower than would be otherwise.

According to a next box 870, a discard decision is made for the specific packet, in

15   accordance with the known type of speech. If optional box 860 has been executed, the discard decision is made in accordance with the set discarding probability.

According to a next box 880, it is inquired if the discard decision is to drop the packet. If not, then execution returns to box 810.

If yes, then according to a next box 890, the packet is deleted from the buffer

20   without being retransmitted. Then execution returns to box 810.

A person skilled in the art will be able to practice the present invention in view of the description present in this document, which is to be taken as a whole. Numerous details have been set forth in order to provide a more thorough understanding of the invention. In other instances, well-known features have not been described in detail in

25   order not to obscure unnecessarily the invention. For example, the teachings of the invention may also be applied to conferencing of more than two calls, using voice and/or video, etc.

While the invention has been disclosed in its preferred form, the specific embodiments as disclosed and illustrated herein are not to be considered in a limiting

30   sense. Indeed, it should be readily apparent to those skilled in the art in view of the present description that the invention may be modified in numerous ways. The inventor

regards the subject matter of the invention to include all combinations and subcombinations of the various elements, features, functions and/or properties disclosed herein.

The following claims define certain combinations and subcombinations, which are regarded as novel and non-obvious. Additional claims for other combinations and subcombinations of features, functions, elements and/or properties may be presented in this or a related document.